

Sequences Associated With Human Iris Pigmentation

Tony Frudakis,¹ Matthew Thomas, Zach Gaskin, K. Venkateswarlu,
K. Suresh Chandra, Siva Gijupalli, Sitaram Gunturi,
Sivamani Natrajan, Viswanathan K. Ponnuswamy
and K. N. Ponnuswamy

DNAPrint Genomics, Sarasota, Florida 34236

Manuscript received March 28, 2003

Accepted for publication August 20, 2003

ABSTRACT

To determine whether and how common polymorphisms are associated with natural distributions of iris colors, we surveyed 851 individuals of mainly European descent at 335 SNP loci in 13 pigmentation genes and 419 other SNPs distributed throughout the genome and known or thought to be informative for certain elements of population structure. We identified numerous SNPs, haplotypes, and diplotypes (diploid pairs of haplotypes) within the *OCA2*, *MYO5A*, *TYRP1*, *AIM*, *DCT*, and *TYR* genes and the *CYP1A2*-15q22-ter, *CYP1B1*-2p21, *CYP2C8*-10q23, *CYP2C9*-10q24, and *MAOA*-Xp11.4 regions as significantly associated with iris colors. Half of the associated SNPs were located on chromosome 15, which corresponds with results that others have previously obtained from linkage analysis. We identified 5 additional genes (*ASIP*, *MC1R*, *POMC*, and *SILV*) and one additional region (*GSTT2*-22q11.23) with haplotype and/or diplotypes, but not individual SNP alleles associated with iris colors. For most of the genes, multilocus gene-wise genotype sequences were more strongly associated with iris colors than were haplotypes or SNP alleles. Diploypes for these genes explain 15% of iris color variation. Apart from representing the first comprehensive candidate gene study for variable iris pigmentation and constituting a first step toward developing a classification model for the inference of iris color from DNA, our results suggest that cryptic population structure might serve as a leverage tool for complex trait gene mapping if genomes are screened with the appropriate ancestry informative markers.

IRIS pigmentation is a complex genetic trait that has long interested geneticists, anthropologists, and the public at large. However, it is yet to be completely understood. Eumelanin (brown pigment) is a light-absorbing polymer synthesized in specialized melanocyte lysosomes called melanosomes. Within the melanosomes, the tyrosinase (TYR) gene product catalyzes the rate-limiting hydroxylation of tyrosine to 3, 4-dihydroxyphenylalanine (DOPA), and the resulting product is oxidized to DOPAquinone to form the precursor for eumelanin synthesis. Although TYR is centrally important for this process, pigmentation in animals is not simply a Mendelian function of TYR or of any other single protein product or gene sequence. In fact, study of the transmission genetics for pigmentation traits in humans and various model systems suggests that variable pigmentation is a function of multiple heritable factors whose interactions appear to be quite complex (BRAUER and CHOPRA 1978; BRTO *et al.* 1997; BOX *et al.* 1997, 2001; AKEY *et al.* 2001; STURM *et al.* 2001). For example, unlike human hair color (STURM *et al.* 2001), there appears to be only a minor dominance component for mammalian iris color determination (BRAUER and CHOPRA 1978),

and minimal correlation exists among skin, hair, and iris color within or between individuals of a given population. In contrast, between-population comparisons show good concordance; populations with darker average iris color also tend to exhibit darker average skin tones and hair colors. These observations suggest that the genetic determinants for pigmentation in the various tissues are distinct and that these determinants have been subject to a common set of systematic and evolutionary forces that have shaped their distribution in world populations.

At the cellular level, variable iris color in healthy humans is the result of the differential deposition of melanin pigment granules within a fixed number of stromal melanocytes in the iris (IMESCH *et al.* 1997). The density of granules appears to reach genetically determined levels by early childhood and usually remains constant throughout later life, although a small minority of individuals exhibit changes in color during later stages of life (BRTO *et al.* 1997). Pedigree studies in the mid-1970s suggested that iris color variation is a function of two loci: a single locus responsible for depigmentation of the iris, not affecting skin or hair, and another pleiotropic gene for reduction of pigment in all tissues (BRUES 1975). Most of what we have learned about pigmentation since has been derived from molecular genetics studies of rare pigmentation defects in humans and

¹Corresponding author: DNAPrint Genomics, 900 Coconut Ave., Sarasota, FL 34236. E-mail: tfrudakis@dnaprint.com

model systems such as mouse and *Drosophila*. For example, dissection of the oculocutaneous albinism (OCA) trait in humans has shown that many pigmentation defects are due to lesions in the *TYR* gene, resulting in their designation as *TYR*-negative OCAs (Oetting and King 1991, 1992, 1993, 1999; see albinism database at <http://www.cbc.umn.edu/tad/>). *TYR* catalyzes the rate-limiting step of melanin biosynthesis and the degree to which human irises are pigmented correlates well with the amplitude of *TYR* message levels (Lindsey et al. 2001). Nonetheless, the complexity of OCA phenotypes illustrates that *TYR* is not the only gene involved in iris pigmentation (Lee et al. 1994). Although most *TYR*-negative OCA patients are completely depigmented, dark-iris albino mice (C44H) and their human type IB oculocutaneous counterparts exhibit a lack of pigment in all tissues except for the iris (Schmidt and Beermann 1994). Study of a number of other *TYR*-positive OCA phenotypes has shown that, in addition to *TYR*, the oculocutaneous 2 (*OCA2*; Hamabe et al. 1991; Gardner et al. 1992; Durham-Pierre et al. 1994, 1996), tyrosinase-like protein (*TYRP1*; Abbott et al. 1991; Chintamaneni et al. 1991; Boissy et al. 1996), melanocortin receptor (*MC1R*; Robbins et al. 1993; Smith et al. 1998; Flanagan et al. 2000), and adaption 3B (*AP3B*) loci (Ooi et al. 1997), and other genes (reviewed by Sturm et al. 2001) are necessary for normal human iris pigmentation. Each of these genes is part of the main (*TYR*) human pigmentation pathway. In *Drosophila*, iris pigmentation defects have been ascribed to mutations in >85 loci contributing to a variety of cellular processes in melanocytes (Ooi et al. 1997; Lloyd et al. 1998), but mouse studies have suggested that ~14 genes preferentially affect pigmentation in vertebrates (reviewed in Sturm et al. 2001) and that disparate regions of the *TYR* and other OCA genes are functionally distinct for determining the pigmentation in different tissues. Human pigmentation genes break out into several biochemical pathways, including those for tyrosinase enzyme complex formation on the inner surface of the melanosome, hormonal and environmental regulation, melanoblast migration and differentiation, the intracellular routing of new proteins into the melanosome, and the proper transportation of the melanosomes from the body of the cell into the dendritic arms toward the keratinocytes. Nonetheless, the study of human OCA mutants suggests that the number of highly penetrant phenotypically active pigmentation loci is surprisingly small.

Although research on pigment mutants has made clear that a small subset of genes is largely responsible for catastrophic pigmentation defects in mice and humans, it remains unclear whether or how common single-nucleotide polymorphisms (SNPs) in these genes contribute toward (or are linked to) natural variation in human iris color. A brown-iris locus was localized to an interval containing the *OCA2* and *MYO3A* genes (Eiberg and Mohr 1996), and specific polymorphisms

in the *MC1R* gene have been shown to be associated with red hair and blue iris color in relatively isolated populations (Robbins et al. 1993; Valverde et al. 1995; Koppula et al. 1997; Smith et al. 1998; Schiöth et al. 1999; Flanagan et al. 2000). An *ASIP* polymorphism is reported to be associated with both brown iris and hair color (Kanetsky et al. 2002). However, the penetrance of each of these alleles appears to be low and, in general, they appear to explain but a very small amount of the overall variation in iris colors within the human population (Spritz 1995). However, single-gene studies have not provided a sound basis for understanding the complex genetics of human iris color. Because most human traits have complex genetic origins, wherein the whole is often greater than the sum of its parts, innovative genomics-based study designs and analytical methods for screening genetic data *in silico* that are respectful of genetic complexity are needed—for example, the multifactorial and/or phase-known components of dominance and epistatic genetic variance. The first step, however, is to define the complement of loci that on a sequence level explain variance in trait value and, of these, those that do so in a marginal or penetrant sense will be the easiest to find. It is toward this goal that we have performed the present study.

We have applied a nonsystematic, hypothesis-driven genome-screening approach to identify various SNPs, haplotypes, and diploypes marginally (i.e., independently) associated with iris color variation. Our results show that a surprisingly large number of polymorphisms in a large number of genes are associated with iris colors, suggesting that the genetics of iris color pigmentation are quite complex. The sequences we have identified constitute a good first step toward developing a classifier model for the inference of iris colors from DNA, and the nature of some of these as markers of population structure might have implications for the design of other complex trait gene-mapping studies.

MATERIALS AND METHODS

Specimens: Specimens for resequencing were obtained from the Coriell Institute in Camden, New Jersey. Specimens for genotyping were of self-reported European descent, of different age, sex, hair, iris, and skin shades and they were collected using informed consent guidelines under Investigational Review Board guidance. Donors checked a box for blue, green, hazel, brown, black, or unknown/not clear iris colors, and each had the opportunity to identify whether iris color had changed over the course of their lives or whether the color of each iris was different. Individuals for whom iris color was ambiguous or had changed over the course of life were eliminated from the analysis. In addition, for 103 of the subjects, iris colors were reported using a number from 1 to 11 as well, where 1 is the darkest brown/black and 11 is the lightest blue, identified using a color placard. For these subjects, we obtained digital photographs of the right iris, where subjects peered into a box at one end at the camera at the other end to standardize lighting conditions and distance and from which a judge assigned the sample to a color group.

Comparing the results of the two methods of classification, 86 of the classifications matched. Of the 17 that did not, 6 were brown/hazel, 7 were green/hazel, and 4 were blue/green discrepancies although none were gross discrepancies such as brown/green, brown/blue, or hazel/blue. Although such an error is tolerable for identifying sequences marginally associated with iris colors, the use of the sequences described herein for iris color classification would therefore likely require digitally quantified iris colors (which we have begun to accumulate and will present elsewhere).

SNP discovery: We obtained candidate SNPs from the National Center for Biotechnology Information (NCBI) Single Nucleotide Polymorphism Database (dbSNP), which generally provided more candidate SNPs than were possible to genotype. We focused on human pigmentation and xenobiotic metabolism genes, selected on the basis of their gene identities, not their chromosomal position. For some genes, the number of SNPs in the database was low and/or some of the SNPs were strongly associated with iris colors, warranting a deeper investigation. For these genes we performed resequencing and of the genes discussed in this article, 113 SNPs were discovered in *CYP1A2* (7 gene regions, 5 amplicons, 10 SNPs found), *CYP2C8* (9 gene regions, 8 amplicons, 15 SNPs found), *CYP2C9* (9 gene regions, 8 amplicons, 24 SNPs found), *OCA2* (16 gene regions, 15 amplicons, 40 SNPs found), *TYR* (5 gene regions, 5 amplicons, 10 SNPs found), and *TYRP1* (7 gene regions, 6 amplicons, 14 SNPs found). Resequencing for these genes was performed by amplifying the proximal promoter (average 700 bp upstream of transcription start site), each exon (average size 1400 bp), the 5' and 3' ends of each intron (including the intron-exon junctions, average size ~100 bp), and 3' untranslated region (UTR; average size 700 bp) sequences from a multi-ethnic panel of 672 individuals (450 individuals from the Coriell Institute's DNA Polymorphism Discovery Resource, 96 additional European Americans, 96 African Americans, 10 Pacific Islanders, 10 Japanese, and 10 Chinese; these 672 individuals represented a set of samples separate from that used for the association study described herein). PCR amplification was accomplished using *pfu* Turbo polymerase according to the manufacturer's guidelines (Stratagene, La Jolla, CA). We developed a program (T. FRUDAKIS, M. THOMAS, Z. GASKIN, K. VENKATESWARLU, K. SURESH CHANDRA, S. GINJUPALLI, S. GUNTURI, S. NATRAJAN, V. K. PONNUSWAMY and K. N. PONNUSWAMY, unpublished results) to design resequencing primers in a manner respectful of homologous sequences in the genome, to ensure that we did not coamplify pseudogenes or amplify from within repeats. BLAST searches confirmed the specificity of all primers used. Amplification products were subcloned into the pTOPO (Invitrogen, San Diego) sequencing vector and 96 insert-positive colonies were grown for plasmid DNA isolation (the use of 670 individuals for the amplification step reduced the likelihood of an individual contributing more than once to this subset of 96 selected). We sequenced with an ABI3700 using PE Applied Biosystems BDT chemistry and we deposited the sequences into a commercial relational database system (iFINCH, Geospiza, Seattle). PHRED-qualified sequences were imported into the CLUSTAL X alignment program and the output of this was used with a second program that we developed (T. FRUDAKIS, M. THOMAS, Z. GASKIN, K. VENKATESWARLU, K. SURESH CHANDRA, S. GINJUPALLI, S. GUNTURI, S. NATRAJAN, V. K. PONNUSWAMY and K. N. PONNUSWAMY, unpublished results) to identify quality-validated discrepancies between sequences. We selected those for which at least two instances of PHRED identified variants that scored ≥ 24 , and each of these SNPs discovered through resequencing were used for genotyping.

Genotyping: For most of the SNPs, a first round of PCR was performed on the samples using the high-fidelity DNA

polymerase *pfu* Turbo and the appropriate resequencing primers. Representatives of the resulting PCR products were checked on an agarose gel, and first-round PCR product was diluted and then used as template for a second round of PCR. The two rounds were necessary due to the fact that many of the genes we queried were members of gene families, the SNPs resided in regions of sequence homology, and our genotyping platform required short (~100 bp) amplicons. For those remaining, only a single round of PCR was performed. Genotyping was performed for individual DNA specimens using a single base primer extension protocol and an SNPstream 25K/ultra-high throughput (UHT) instrument (Beckman Coulter, Fullerton, CA, and Orchid Biosystems, Princeton, NJ). Genotypes were subject to several quality controls: two scientists independently pass/fail inspected the calls, requiring an overall UHT signal intensity >1000 for >95% of genotypes and clear signal differential between the averages for each genotype class (i.e., clear genotype clustering in two-dimensional space using the UHT analysis software).

Statistical methods: To test the departures from independence in allelic state within and between loci, we used the exact test, described in ZAYKIN *et al.* (1995). Haplotypes were inferred using the STEPHENS *et al.* (2001) haplotype reconstruction method. To determine the extent to which extant iris color variation could be explained by various models, we calculated R^2 values for SNPs, haplotypes, and multilocus genotype data by first assigning the phenotypic value for blue eye color as 1, green eye color as 2, hazel eye color as 3, and brown eye color as 4. Biogeographical ancestry admixture proportions were determined using the methods of HANIS *et al.* (1986) and SHRIVER *et al.* (2003) within the context of a software program we developed for this purpose, which will be presented elsewhere (T. FRUDAKIS, Z. GASKIN, M. THOMAS, V. PONNUSWAMY, K. VENKATESWARLU, S. GUNJUPULLI, C. BONILLA, E. PARRA, and M. SHRIVER, personal communication). For R^2 computation, we used the following function: $\text{Adj-}R^2 = 1 - [n/(n-p)](1-R^2)$, where n is the model degrees of freedom and $n-p$ is the error degrees of freedom. To correct for multiple tests, we used the empirical Bayes adjustments for multiple results method described by STEINLAND *et al.* (2000). Linkage disequilibrium (LD) for pairs of SNPs within a gene was determined using the Zaykin exact test and a cutoff value of $|D'| \geq 0.05$ (P value < 0.05; ZAYKIN *et al.* 1995).

RESULTS

To identify SNP loci associated with variable human pigmentation, we genotyped for 754 SNPs: 335 SNPs within pigmentation genes (*AP3B1*, *ASIP*, *DCT*, *MC1R*, *OCA2*, *SILV*, *TYR*, *TYRP1*, *MYO5A*, *POMC*, *AIM*, *AP3D1*, and *RAB*, Table 1), and 419 other SNPs distributed throughout the genome. Alleles for these latter SNPs were known to be informative for certain elements of population structure; 73 were selected from a screen of the human genome because they were exceptional ancestry informative markers (AIMs, based on high δ values) for Indo-European, sub-Saharan African, Native American, and East Asian biogeographical ancestry (BGA; SHRIVER *et al.* 2003; T. FRUDAKIS, Z. GASKIN, M. THOMAS, V. PONNUSWAMY, K. VENKATESWARLU, S. GUNJUPULLI, C. BONILLA, E. PARRA and M. SHRIVER, unpublished observations). The rest were found in or around xenobiotic metabolism genes, which we have previously shown exhibit dramatic sequence variation

TABLE 1
Candidate genes tested for sequence associations with human iris pigmentation

Gene	Name	Homology/model phenotype*	References
<i>AP3B1</i>	Adaptor-related protein complex 3, β -1 subunit	Mouse "pearl" human HPS2	BALKEMA <i>et al.</i> (1983)
<i>ASIP</i>	Agouti signaling protein	Mouse "agouti"	KWON <i>et al.</i> (1994) KLEBIG <i>et al.</i> (1995)
<i>DCT</i>	Dopachrome tautomerase	TYR-related protein 2 mouse "slaty"	KWON (1993) JACKSON <i>et al.</i> (1992)
<i>MC1R</i>	Melanocortin 1 receptor	Mouse "extension" (c ADP)	ROBBINS <i>et al.</i> (1993)
<i>OCA2</i>	Oculocutaneous albinism II	Mouse pink-eyed dilution (p)	LYON <i>et al.</i> (1992) GARDNER <i>et al.</i> (1992)
<i>SILV</i>	Silver homologue	Mouse "silver" (si)	KWON <i>et al.</i> (1991)
<i>TYR</i>	Tyrosinase	Mouse "albino" (c), Himalayan	BARTON <i>et al.</i> (1988) KING <i>et al.</i> (1989) KWON <i>et al.</i> (1989)
<i>TYRP1</i>	Tyrosinase related protein 1	Mouse "brown" (b)	JACKSON (1988)
<i>MYO5A</i>	Myosin VA	Mouse "dilute" (d)	COPELAND <i>et al.</i> (1983) STROBEL <i>et al.</i> (1990)
<i>POMC</i>	Proopiomelanocortin	Mouse Pomc1	KRUDE <i>et al.</i> (1998)
<i>AIM (MATP or AIM-1)</i>	Membrane associated transporter protein	Mouse "underwhite" (uw)	NEWTON <i>et al.</i> (2001)
<i>AP3D1</i>	Adaptor-related protein complex 3, Δ -1 subunit	Mouse "mocha" (mh)	None
<i>RAB</i>	RAB27A oncogene	Mouse "ashen" (ash)	WILSON <i>et al.</i> (2000)

* Name of mutant is in parentheses.

as a function of BGA (FRUDAKIS *et al.* 2003). Genotypes for these 754 candidate SNPs were scored for 851 European-derived individuals of self-reported iris colors (292 blue, 100 green, 186 hazel, and 273 brown). Before screening these genotypes for association with iris colors, we used the 73 nonxenobiotic metabolism AIMS to determine BGA admixture proportions for each sample and we tested for correlation between BGA admixture and iris colors. This test showed that each of our 851 Caucasian samples was of majority Indo-European BGA, and although 58% of the samples were of significant (>4%) non-Indo-European BGA admixture, there was no correlation among low levels (<33%) of East Asian, sub-Saharan African, or Native American admixture and iris colors. For more extensively admixed individuals, we observed no correlation between higher levels (>33% but <50%) of Native American admixture and iris colors, although there was a weak association between higher levels of East Asian and sub-Saharan African admixture and darker iris colors (data not shown).

It was unclear from the outset whether we would have better success considering iris color in terms of four colors (blue, green, hazel, and brown) or in terms of groups of colors. One method of grouping colors is light = blue + green and dark = hazel + brown, and this grouping would seem to more clearly distinguish individuals with respect to the detectable level of eumelanin (brown pigment). Given that our iris color data were self-reported, partitioning the sample into brown and not brown, or blue and not blue, could provide

greater power to detect significant associations, particularly for alleles associated with blue or brown irises. To take advantage of each of these four methods, we considered all of them when screening SNPs for associations; we calculated the δ value, chi square, and exact test *P* values for (a) all four colors, (b) shades, using light (blue and green) vs. dark (hazel and brown), (c) blue vs. brown, and (d) brown vs. not brown (blue, green, and hazel) groupings. We fixed significance levels at 5%, and the alleles of 20 SNPs were found to be associated with specific iris colors, 19 with iris color shades, 19 with blue/brown color comparisons, and 18 using the brown/not brown comparison. The overlap among these SNP sets was high but not perfect. In all, 27 SNPs were significantly associated with iris pigmentation using at least one of the four criteria, and we refer to these as "marginally" associated. When multiple simultaneous hypotheses are tested at set *P* values, there is the possibility of enhanced type I error, so we used the correction procedure of STEENLAND *et al.* (2000) with adjusted residuals to compensate for this risk. We found that most of the associations were still significant after this correction (those with asterisks in Table 2), and since the analysis was conducted using adjusted residuals, some new associations were observed (*i.e.*, *MAOA* marker 2 had a chi-square *P* value of 0.24 but was associated using the corrected testing procedure; Table 2). Most of the marginally associated SNPs were found within the pigmentation genes *OCA2* (*n* = 10), *TYRP1* (*n* = 4), *AIM* (*n* = 3), *MYO5A* (*n* = 2), and *DCT* (*n* =

TABLE 2
SNPs marginally (independently) associated with iris pigmentation and SNPs associated only within the context of haplotypes and/or diplotypes

Gene	Haplotype order ^a	Sequence	F_{CA} (minor) ^b	HWE- P^c	Colors P	Shade P	Brown/blue P	Brown/other P	Pigment history
AIM (MATP) 5p13.3	1	rs35391	0.03-T,C	0.04	0.01*	0.02	<0.01*	0.02	None
	2	rs40132	0.02-G,A	0.32	0.02	0.01	<0.01	0.02	None
	3	rs26722	0.02-T,C	0.29	0.08	0.04*	0.01	0.08	None
ASIP 20q11.2-q12	1	rs2424987	0.12-G,A	0.13	0.36	0.51	0.84	0.50	Brown irises, KANETSKY <i>et al.</i> (2002)
	2	rs2424984	0.13-C,T	0.14	0.81	0.73	0.86	0.75	None
	3	rs819135	0.42-G,A	0.50	0.92	0.51	0.68	0.82	None
CYP1A2 15q22-ter	1	AY392136	0.45-G,C	0.98	0.09	0.71	0.11	0.03*	None
CYP1B1 2p21	1	rs162560	0.24-T,C	0.01	0.11*	0.02*	0.11	0.46	None
	2	rs1056837	0.44-T,C	0.13	0.84	0.50	0.46	0.71	None
CYP2C8 10q23	1	rs1926705	0.32-G,T	0.01	0.10	0.30	0.04*	0.03*	None
	2	rs1341164	0.30-G,A	0.70	0.45	0.21	0.14	0.11	None
	3	AY392132	0.12-G,A	0.80	0.15	0.91	0.95	0.84	None
CYP2C9 10q24	1	AY392131	0.32-C,T	0.70	0.13	0.07*	0.11	0.05*	None
DCT 13q32	1	rs1407995	0.18-T,C	0.17	<0.01*	0.35	0.64	0.14	None
	2	rs1325611	0.18-C,T	0.05	<0.01*	0.50	0.91	0.08	None
	3	rs2892681	0.34-C,G	0.42	0.37	0.44	0.56	0.75	None
	4	rs1028806	0.20-G,A	0.94	0.45	0.37	0.66	0.40	None
	5	rs2031527	0.21-T,C	0.81	0.67	0.99	0.85	0.71	None
	6	rs2296498	0.11-G,A	0.37	0.88	0.51	0.66	0.91	None
GSTT2 22q11.23	1	rs2140186	0.46-G,A	0.18	0.40	0.37	0.92	0.85	None
	2	rs2140196	0.49-G,A	0.06	0.55	0.33	0.74	0.63	None
MAOA Xp11.4-11.3	1	rs979605	0.28-T,C	<0.01	0.05*	0.01*	0.04*	0.27	None
	2	rs2072743	0.33-A,G	<0.01	0.24*	0.05*	0.09	0.28	None
MC1R 16q24.3	1	rs1805007	0.09-T,C	0.14	0.18	0.31	0.67	0.29	Red hair/blue irises, multiple authors
	2	rs1805008	0.07-T,C	0.37	0.69	0.25	0.45	0.64	Red hair/blue irises, multiple authors
	3	rs2228479	0.08-T,C	<0.01	0.84	0.58	0.94	0.83	Red hair/blue irises, multiple authors
MYO5A 15q21	1	rs1724630	0.20-G,C	0.76	0.02*	0.05*	0.22	0.74	None
	2	rs1724631	0.24-T,G	<0.01	0.05*	0.32	0.55	0.41	None
	3	rs2290332	0.13-G,A	0.66	0.08*	0.24	0.38	0.64	None
	4	rs752864	0.38-C,T	0.80	0.22*	0.08	0.11	0.45	None
	5	rs1724639	0.38-T,C	0.51	0.17*	0.14	0.20	0.82	None
	6	rs935892	0.40-G,A	0.54	0.25	0.49	0.40	0.87	None
	7	rs1615235	0.22-A,G	0.38	0.19	0.09	0.31	0.92	None
	8	rs2242057	0.35-T,C	0.78	0.28	0.13	0.29	0.85	None
	9	rs2899488	0.16-T,C	0.71	0.89	0.69	0.99	0.78	None
	10	rs1869126	0.16-T,C	0.65	0.93	0.53	0.70	0.79	None
OCA2 15q11.2-q12	1	rs749846	0.15-T,G	0.34	<0.01*	<0.01*	<0.01*	<0.01*	None
	2	AY392135	0.18-A,G	0.08	<0.01*	<0.01*	<0.01*	<0.01*	None
	3	AY392134	0.28-C,A	0.90	<0.01*	<0.01*	<0.01*	<0.01*	None
	4	rs1900758	0.30-G,A	0.81	<0.01*	<0.01*	<0.01*	<0.01*	None
	5	rs1037208	0.18-C,A	0.13	<0.01*	<0.01*	<0.01*	<0.01*	None
	6	rs1800411	0.27-C,T	0.78	<0.01*	<0.01*	<0.01*	0.02*	None
	7	rs1800404	0.21-G,A	0.95	<0.01*	<0.01*	<0.01*	0.02*	None
	8	AY392133	0.04-C,G	0.65	0.02	0.07	0.01*	<0.01*	None
	9	rs1800401	0.04-T,C	0.63	0.03	0.02*	<0.01*	<0.01*	Brown irises, REBBECK <i>et al.</i> (2003)
	10	rs2044627	0.38-G,A	0.19	0.05*	0.29	0.02*	0.03*	None
	11	rs1448483	0.04-G,A	<0.01	0.13	0.06	0.31	0.25	None
	12	rs737051	0.29-G,A	0.16	0.37	0.39	0.16	0.08	None
	13	rs1800410	0.11-G,A	0.96	0.46	0.5	0.25	0.12	None

(continued)

TABLE 2
(Continued)

Gene	Haplotype order ^a	Sequence	F_{CA} (minor) ^b	HWE- P ^c	Colors P	Shade P	Brown/ blue P	Brown/ other P	Pigment history
POMC 2p23	1	rs934778	0.30-C,T	0.11	0.87	0.77	0.91	0.67	None
SILV 12q13-q14	1	rs1052206	0.18-C,T	0.08	0.43	0.68	0.22	0.18	None
	2	rs1052165	0.29-T,C	0.53	0.42	0.76	0.92	0.37	None
TYR 11q14-q21	1	rs1827430	0.45-G,A	0.41	0.06	0.37	0.52	0.65	None
	2	rs1042602	0.36-A,C	0.06	0.64	0.55	0.72	0.45	None
	3	rs1851992	0.42-G,A	0.19	0.92	0.97	0.76	0.58	None
TYRP1 9p23	1	rs2733832	0.41-C,T	0.45	0.02*	<0.01*	<0.01*	<0.01*	None
	2	rs2075508	0.06-C,T	0.54	0.06	<0.01*	0.04	0.08	None
	3	rs683	0.35-G,T	0.19	0.01	<0.01*	<0.01*	<0.01*	None
	4	rs2762464	0.35-A,T	0.85	0.03	0.04*	0.01*	<0.01*	None
	5	AY395737	0.06-T,C	0.89	0.39	0.09	0.13	0.17	None
	6	AY395736	0.06-T,G	0.92	0.62	0.20	0.24	0.43	None

Asterisks represent P values that remained significant after the correction for multiple tests and P values in italic are those that were statistically significant ($P \leq 0.05$).

^a Haplotype order refers to the order of the SNPs in the haplotypes shown in Table 4 and described in the text.

^b Frequency of the minor allele and the major and minor allele nucleotide.

^c Hardy-Weinberg equilibrium P value, where a value <0.05 indicates that the alleles are not in equilibrium.

2) although some associations were found within non-pigmentation genes such as CYP2C8 at 10q23, CYP2C9 at 10q24, CYP1B1 at 2p21, and MAOA at Xp11.3. No significant SNP associations within the pigmentation genes *SILV*, *MC1R*, *ASIP*, *POMC*, *RAB*, or *TYR* were found, although *TYR* had one SNP with a $P = 0.06$. The most strongly associated of the marginally associated SNPs were from the *OCA2*, *TYRP1*, and *AIM* genes, in order of the strength of association, which is the same order as that provided using the number of marginally associated SNPs, rather than their strength.

Since most of the SNPs identified from this approach localized to discrete genes or chromosomal regions, we grouped all of the SNPs from each locus and tested inferred haplotypes for association with iris colors using contingency analysis. We did not confine this higher-order analysis to those genes with marginal SNP associations, but we grouped all of the high-frequency SNPs tested for each gene. For each gene, we inferred haplotypes and used contingency analyses to determine which haplotypes were statistically associated with iris colors. From the chi-square and adjusted residuals, we found 43 haplotypes for 16 different loci to be either positively (agonist) or negatively (antagonist) associated with iris colors (Table 3). The strongest associations were observed for genes with SNPs that were marginally associated (Table 2) and most of the genes with marginal SNP associations had haplotypes and diplotypes (sometimes referred to as multilocus gene-wise genotypes or diploid pairs of haplotypes) positively (agonist) or negatively (antagonist) associated with at least one iris color (Table 3). A few of the genes/regions not harboring a marginally associated SNP had haplotypes and diplotypes positively and/or negatively associated with iris colors (*ASIP*

gene, 1 haplotype; *MC1R* gene, 2 haplotypes; Tables 2 and 3). In other words, their SNPs were associated with iris colors only within the context of gene haplotypes or diplotypes. For some, associations with iris colors were found only within the context of diplotypes, but not at the level of the SNPs or the haplotype (i.e., *SILV* and *GSTT2* genes located at 22q11.23). At the level of the haplotype, each gene or region had unique numbers and types of associations. For example, *OCA2*, *AIM*, *DCT*, and *TYRP1* harbored haplotypes both positively associated with blue irises and negatively associated with brown irises (*OCA2* haplotypes 1, 37, 38, 42; *AIM* haplotype 1; *DCT* haplotype 2; and *TYRP1* haplotype 1; Table 3). Others genes such as *AIM*, *OCA2*, and *TYRP1* harbored haplotypes positively associated with brown but negatively associated with blue color (*AIM* haplotype 2; *OCA2* haplotypes 2, 4, 45, 47; *TYRP1* haplotype 4; Table 3) while others, such as the *MYO5A*, *OCA2*, *TYRP1*, and *CYP2C8* genes located at 10q23, harbored haplotypes positively associated with one color but not negatively associated with any other color (*MYO5A* haplotype 5 and haplotype 10, *OCA2* haplotype 19, *TYRP1* haplotype 3, and *CYP2C8* haplotype 1; Table 3). The *MC1R* gene harbored haplotypes associated only with green color in our sample and the *POMC* gene harbored a single SNP with genotypes weakly associated with iris colors (no significant haplotypes or diplotypes were found). Overall, the diversity of haplotypes associated with brown irises was similar to that of haplotypes associated with blue irises. Most of the haplotypes were even more dramatically associated with iris colors in a multiracial sample (data not shown), because many of the SNPs comprising them are good AIMS and variants associated with darker iris colors were enriched in those ancestral

TABLE 3
The common haplotypes and diplotypes for the 16 iris color genes discussed in the text

Sequence ID	Haplotypes of 16 genes*	Agonist [†] chi-square <i>P</i>	Agonist [†] color	Antagonist [†] chi-square <i>P</i>	Antagonist [†] color	Count [‡]
AIM (MATP) 5p13.3						
1	C A C	0.0010	Blue	0.004	Brown	(1641)
2	T G T	0.048	Brown	0.003	Blue	(33)
3	T A C	—	—	—	—	(23)
ASIP 20q11.2-q12						
1	A T A	—	—	—	—	(979)
2	A T G	—	—	—	—	(508)
3	G C G	—	—	—	—	(196)
4	A C A	0.017	Hazel	—	—	(13)
DCT 13q32						
1	C T G A C A	—	—	—	—	(625)
2	C T C A C A	0.014	Hazel	0.028	Blue	(242)
		0.016	Blue	0.021	Brown	
3	T C G A C A	0.048	Green	0.003	Hazel	(281)
4	C T C G T A	—	—	—	—	(320)
5	C T G A C G	—	—	—	—	(179)
MCIR 16q24.3						
1	T C C	0.016	Green	—	—	(152)
2	C C C	—	—	0.026	Green	(1294)
3	C C T	—	—	—	—	(143)
4	C T C	—	—	—	—	(113)
MYO5A 15q21						
1	C G A T C G G C C C	0.000	Green	—	—	(51)
2	C G A T C A G C C C	—	—	—	—	(858)
4	G T G C T G A T C C	—	—	0.008	Blue	(163)
5	C G A T C A A C C C	0.017	Blue	—	—	(40)
6	G T A C T G A T C C	—	—	—	—	(117)
8	C G A C T G G T T T	—	—	—	—	(165)
10	C G A C C A G C C C	0.003	Brown	—	—	(40)
13	C T A C T G G T T T	—	—	—	—	(71)
14	C G A T T A G C C C	—	—	0.027	Blue	(40)
16	C G G C C A A T C C	—	—	—	—	(19)
OCA2 15q11.2-q12						
1	G G G G A C G G C A A A G	0.002	Blue	0.01	Brown	(44)
2	G A G G C C G G C A A G A	0.018	Hazel	0.031	Blue	(43)
		0.022	Brown	—	—	
3	G A G G C C A G C A A G A	0.042	Green	0.026	Blue	(21)
4	G A G G C C G G C A A A A	0.024	Brown	0.014	Blue	(89)
7	G A G G C C A G C A A A A	—	—	—	—	(13)
15	T G G G A C G C T A A A G	—	—	—	—	(17)
19	G A G G C C A G C G G A G A	0.019	Brown	—	—	(23)
22	G A G G C C G G C G A A A	0.036	Green	0.012	Blue	(15)
25	T A A G C C A G C G A A A	—	—	—	—	(13)
37	G G A A A T A G C A A A A	<0.001	Blue	<0.001	Brown	(508)
38	G G A A A T A G C G A A A	0.025	Blue	0.011	Brown	(200)
39	G G A A A T A G C A A G A	—	—	0.039	Brown	(71)
41	G G A A A T A G C A G A A	—	—	—	—	(19)
42	G G A A A T A G C G A G A	0.029	Blue	0.021	Brown	(174)
45	T G A A A T A G C G A A A	<0.001	Brown	0.007	Blue	(65)
47	T G A A A T A G C G A G A	0.003	Brown	0.006	Hazel	(22)
		—	—	0.036	Blue	
48	T G A A A T A G C A A A A	0.001	Brown	0.015	Green	(35)
57	T G A A A T A G C A A G A	—	—	—	—	(30)

(continued)

TABLE 3
(Continued)

Sequence ID	Haplotypes of 16 genes*	Agonist [†] chi-square <i>P</i>	Agonist [†] color	Antagonist [†] chi-square <i>P</i>	Antagonist [†] color	Count [‡]
POMC 2p23.3						
1	T	—	—	—	—	(1187)
2	C	—	—	—	—	(515)
2/2	C/C	0.043	Blue	—	—	
SILV 12q13-q14						
1	T C	—	—	—	—	(913)
2	C C	—	—	—	—	(296)
3	T T	—	—	—	—	(490)
1/1	TC/TC	—	—	0.035	Blue	
TYR 11q14-q21						
1	G C G	—	—	0.01	Green	(184)
2	A A G	—	—	—	—	(257)
3	G C A	—	—	—	—	(467)
4	A A A	—	—	—	—	(231)
5	A C A	—	—	—	—	(254)
6	A C G	—	—	—	—	(189)
7	G A G	—	—	—	—	(83)
8	G A A	—	—	0.023	Blue	(37)
TYRP1 9p23						
1	T T T T C G	0.007	Blue	<0.001	Brown	(968)
2	C T T T T T	—	—	—	—	(86)
3	C T G A C G	0.006	Brown	—	—	(454)
4	C C G A C G	0.046	Brown	0.029	Blue	(98)
5	T T G T C G	—	—	—	—	(23)
7	C T T A C G	—	—	—	—	(24)
CYP1A2-15q22-ter						
1	G	0.015	Brown	0.023	Hazel	(769)
2	C	0.023	Hazel	0.015	Brown	(933)
CYP1B1-2p21						
1	C C	—	—	—	—	(957)
2	T T	0.027	Hazel	0.025	Blue	(403)
3	C T	—	—	—	—	(337)
CYP2C8-10q23						
1	C A A	0.007	Brown	—	—	(539)
2	T A G	—	—	—	—	(201)
3	T G A	—	—	0.039	Brown	(513)
4	T A A	—	—	—	—	(439)
CYP2C9-10q24						
1	T	0.023	Brown	0.039	Green	(1325)
2	C	0.039	Green	0.023	Brown	(377)
GSTT2-22q11.23						
1	A G	—	—	—	—	(821)
2	G A	—	—	—	—	(778)
3	A A	—	—	—	—	(99)
	AG/AG	0.040	Green	—	—	(184)
	AG/GA	0.013	Hazel	—	—	(401)
MAOA-Xp11.4-11						
1	C G	—	—	0.027	Blue	(1133)
2	T A	0.006	Blue	0.026	Hazel	(480)
3	C A	—	—	—	—	(87)

* Sequences of the highest order of complexity within a locus found to be associated with iris colors. All of the major sequences (count ≥ 13) for each locus with at least one significantly associated sequence are shown. If no haplotypes or diplootypes for a locus were found to be associated, only the SNP alleles are shown. If no haplotypes were found to be associated for a locus but diplootypes were found to be associated, both the haplotypes and the diplootypes are shown.

[†] Agonist color refers to the color with which the sequence is positively associated. Antagonist color refers to the color with which the sequence is negatively associated. Chi-square *P* value is shown.

[‡] Number of times the haplotype was observed in our sample of 851.

groups of the world that are of darker average iris color (FRUDAKIS *et al.* 2003; data not shown). Most of the SNPs within a gene or region were in LD with others in that gene or region ($|D'| \geq 0.05$); only 32 SNP pairs—in the *MC1R* (1 pair), *OCA2* (27 pairs), *TYR* (2 pairs), and *TYRP1* (2 pairs) genes—were found to be in linkage equilibrium (not shown).

These analyses resulted in the identification of 61 SNPs in 16 genes/chromosomal regions associated with iris colors on one level or another; details for each and whether the SNP is marginally associated or associated within the context of the haplotype and/or diplotypes are shown in Table 2. The minor allele frequency for most of these SNPs was relatively high (average F minor allele = 0.22) and most of them were in Hardy-Weinberg equilibrium (HWE; those for which HWE $P > 0.05$, 28/34; Table 3). Nine were not and of these 2 were of relatively low frequency with weak evidence for disequilibrium (P value close to 0.05). Lack of HWE is usually an indication of a poorly designed genotyping assay, but none of the remaining 7 SNPs exhibited genotyping patterns that we have previously associated with such problems (such as the complete absence of an expected genotype class or all genotypes registering as heterozygotes). Indeed, one of those for which the evidence of lack of HWE was the strongest was validated as a legitimate SNP through direct DNA sequencing (data not shown). The chromosomal distribution of the SNPs that were significantly associated in a marginal sense was found to be independent of the distribution of SNPs actually surveyed, indicating that the associations were not merely a function of SNP sampling and the same was true for the distribution of all the SNPs shown in Table 2 (data not shown). Chromosome 15q harbored the majority (14/27) of the SNPs that were marginally associated with iris colors, and all but one of these 14 were found in two different genes: *OCA2* and *MYO5A* (Table 2). Chromosome 5p had 3 SNPs marginally associated, all in the *AIM* gene, and chromosome 9p had 5 SNPs associated, all in the *TYRP1* gene. Multiple SNPs were identified on chromosome 10q; the *CYP2C8-10p23* region had 1 marginally associated SNP, and the neighboring region, *CYP2C9-10p24*, also had one. As one might expect from the proximity of these two regions, *CYP2C8-CYP2C9* marker pairs were found to be in tight LD with one another ($P < 0.001$ for each possible pair). Multiple SNPs were also identified on chromosome 2; the C/C genotype for the POMC SNP located at 2p23 was associated with blue iris color (Table 3) and a CYP1B1-3p21-region SNP was also marginally associated at the level of iris shade (Table 2), as well as within the context of a 2-SNP haplotype (Table 3). The SNPs between the 2p21 and 2p23 regions were also in LD ($P < 0.01$). Finally, in addition to the *OCA2* (15q11.2–q12) and *MYO5A* (15q21) sequences, a single SNP (15q22–ter) was also implicated on chromosome 15q, but SNPs between each of these three loci were not

found to be in LD (data not shown). SNPs for the *MC1R* (16q24), *SLV* (12q13), and *TYR* (11q) genes and for the *MAOA-Xp11.4–11.3* and *GSTT2-22q11.23* regions were also found to be associated at the level of the haplotype (Tables 3 and 4), although these were the only regions of these chromosomes for which associations were found.

The P values we obtained suggested that diplotypes explained more iris color variation than did haplotypes or individual SNPs. To test this, we performed a corrected ANOVA analysis for our data on each of these three levels. We considered all 61 SNPs in Table 2, their haplotypes in Table 3, and their diplotypes (not shown). Diplotypes explained 15% of the variation, whereas haplotypes explained 13% and SNPs explained only 11% (Table 4) after correcting for the number of variables. The most strongly associated 68 genotypes of the 543 genotypes observed for the 16 genes/regions, on the basis of chi-square-adjusted residuals, explained 13% of the variation (last row in Table 4).

DISCUSSION

From a screen of 754 SNP loci, we have identified 61 that are statistically associated with variable iris pigmentation at one level of intragenic complexity or another. The remaining SNPs had b values and chi-square P values that were not significant on any level of intragenic complexity. Diplotypes for these 61 alleles explained most of the iris color variance in our sample; the lowest amount was explained at the level of the SNP, suggesting an element of intragenic complexity to iris color determination (*i.e.*, dominance). Only about half of the 61 SNPs that we identified were associated with iris colors independently—the others were associated only in the context of haplotypes or diplotypes. Even at this level of complexity, the sequences from no single gene could be used to make reliable iris color inferences, which suggests an element of intergenic complexity (*i.e.*, epistasis) for iris color determination as well. Aside from the fact that many of the SNPs we identified were significant after imposing the Steenland correction for multiple testing, there are three lines of evidence that the SNPs we have identified are not spuriously associated. The first is that for most of the genes for which we identified marginally associated SNPs, multiple such SNPs were identified. In other words, the distribution of SNPs among the various genes tested was not random. Indeed, the associations were observed to be generally stronger for the SNPs in the context of within-gene haplotypes—a result that would not necessarily be obtained for individual SNPs spuriously associated—suggesting that the gene sequences themselves are associated, not merely a spurious polymorphism within each gene. Second, although a roughly equal number of pigmentation and nonpigmentation gene SNPs were tested, of the 34 marginally associated SNPs, 28 of them

TABLE 4
ANOVA-SNP and haplotype data

Source	Model d.f.	Error d.f.	No. of variables	Fvalue	R ² value	Adjusted R ²
SNP data	62	788	62	2.63	0.17	0.11
Haplotype data	212	638	216	1.58	0.34	0.13
Multilocus gene-wise genotype data	543	307	572	1.27	0.69	0.15
Multilocus gene-wise genotype data	68	782	68	2.82	0.2	0.13

(82%) were in pigmentation genes. In other words, the distribution of SNPs among the various gene types was also not random. Third, when applied to a sample including individuals of multiple ancestries, the linear and nonlinear variables from these and the other genes combined performed even better than when applied just to individuals of majority European ancestry (not shown). Since most individuals of non-European or minority European descent exhibit low variability in iris colors (on average of darker shade than individuals of European descent), this improvement may not seem surprising. However, this result would not have necessarily been obtained were we working with SNPs that were not truly associated with iris colors. Although corrections for multiple testing left most of the SNP-level associations intact, a number of the associations we found did not pass the multiple-testing examination, but nonetheless we present them here to avoid possible type II error; the sequences may be weakly associated with iris colors and possibly relevant within a multiple-gene model for classification (*i.e.*, epistasis). For these, it would seem more prudent to eliminate false positives downstream of SNP identification, such as from tests of higher-order association, using various other criteria, such as those described above, or possibly using the utility of the SNP for the generalization of a complex classification model when one is finally described.

Mutations in the pigmentation genes are the primary cause of oculocutaneous albinism so it was natural to expect that common variations in their sequences might explain some of the variance in natural iris colors, and this is in fact what we observed. However, a number of the associations we identified were for SNPs located in other types of genes. The sequences for most of these genes vary significantly as a function of population structure (FRUDAKIS *et al.* 2003) and it is possible that alleles for these SNPs are associated with elements of population structure that correlate with iris colors. Alternatively, the mechanism for the associations could be LD with phenotypically active loci in nearby pigment genes. Indeed, some, but not all, of our nonpigment gene SNPs are found in regions within the vicinity of pigmentation genes; *CYP2C8* and *CYP2C9* are located on chromosome 10 near the *HPS1* and *HPS2* pigmentation genes (which we did not test directly), *CYP1A2* is located at 15q22–ter

on the same arm as *OCA2* and *MYO5A*, *CYP1B1* is located at 2p21 in the vicinity of the *POMC* gene at 2p23, and *MAOA* is located on the same arm of chromosome X (Xp11.4–11.3) as the *OAI* pigmentation gene (which we also did not test directly). The distances between these loci associated with iris colors and "neighboring" pigmentation genes is far greater than the average extent of LD in the genome, and if it is the case that these associations are through LD, it would seem that, again, population structure would need to be invoked as an explanation. The structure behind our results is unlikely to be of a crude (*i.e.*, continental) nature; although two-thirds of our European-American samples were of significant (4%) BGA admixture, few correlations between structure measured on this level and iris colors were observed in this study. Rather, it seems likely that the structure behind our results is of a finer, more "cryptic" nature, such as ethnicity or even within-ethnic-group structure. To an investigator interested in elucidating a biological mechanism, association due to population structure might not seem to be very satisfying, but when classification is the goal rather than the elucidation of a biological mechanism, it would seem to matter little why a marker is associated with a trait. For example, forensic investigators construct physical profiles using surprisingly unscientific means; only in rare cases are eye-witness accounts available, and in certain circumstances these accounts are subjective and unreliable. A battery of genetic tests, of which one for the inference of iris color could be a part, could enable the construction of a more objective and science-based (partial) physical profile from crime-scene DNA, and an investigator using these tests would be less interested in the biological mechanism of the phenotype than in an ability to make an accurate inference of trait value. Of course, identifying markers in LD with phenotypically active loci (or the phenotypically active loci themselves) would provide for more accurate classification (as well as for a better understanding of biological mechanism), but the hunt for these elusive loci in heterogeneous populations is still impractical because LD extends only for a few kilobases and the economics of genome-wide scans in heterogeneous samples with full LD coverage are out of reach for most labs.

Linkage studies have implicated certain pigmentation

genes as specifically relevant for pigmentation phenotypes, and most of the pigmentation gene SNPs that we identified clustered to certain genes such as *OCA2*, *MYO5A*, *TYRP1*, and *AIM*. Further, certain of our results support the previous literature. Most of the SNPs that we identified were on chromosome 15, which EIBERG and MOHR (1996) described from linkage analyses as the primary chromosome for the determination of "brownness." As suggested by these authors, the candidate gene within the interval containing this locus (*BEY2*) is most likely the *OCA2* gene, although the *MYO5A* gene is also present within this interval and, as shown here, associated with iris colors. *OCA2* associations were by far the most significant of any gene or region we tested, while *MYO5A* SNPs were only weakly associated (but haplotypes and diplotypes more strongly). *MYO5A* alleles were not found to be in LD with those of *OCA2*, suggesting that these results were independently obtained and that Eiberg and Mohr's results may have been a reflection of the activity of two separate genes. REBBECK *et al.* (2002) recently described two *OCA2* coding changes associated with darker iris colors. One of these, the Arg305Trp SNP, was one of the 13 *OCA2* SNPs that we found to be strongly associated with iris colors using all four of our color criteria, although its association was only the ninth strongest among the *OCA2* SNPs that we identified and the eleventh strongest among all of the associated SNPs that we identified. The *P* values we obtained for this particular SNP association ($P = 0.01-0.05$, depending on the color criteria) were less significant than those described ($P = 0.002$) by REBBECK *et al.* (2002). In addition, we independently isolated the "red hair/blue iris" SNP alleles described by VALVERDE *et al.* (1995) and KOPPULA *et al.* (1997), suggesting that these sequences are indeed associated with iris pigmentation as suggested by these authors, although we note that the associations described by these authors were with blue irises and at the level of the SNP, while those that we observed were with green irises and apparent only at the level of the haplotypes and diplotypes. We also identified associations in the *ASIP* gene, which supports previous work by KANETSKY *et al.* (2002), although it should be noted that we did not observe this gene association at the level of the SNP as they did; one of the *ASIP* SNPs that we identified (marker 861, Table 2) is the 8818 G-A SNP transversion that they described to be associated with brown iris colors, but from our study the association was with hazel color at the level of the haplotype. Last, we also showed that the associations between *TYR* haplotypes and iris colors were relatively weak, which is not inconsistent with results obtained by many others before us working in the field of oculocutaneous albinism who have failed to find strong associations in smaller samples. Although our results independently verified findings for *OCA2*, *ASIP*, and *MC1R*, they also show that several other pigmentation genes harbor alleles associated with the natural distribution of iris

colors (*TYRP1*, *AIM*, *MYO5A*, and *DCT*). Therefore, it seems that our findings indicate that most of the previous results associating pigmentation gene alleles with iris colors, taken independently, represent merely strokes of a larger, more complex portrait. It is interesting that most of the SNPs that we discovered are noncoding, either silent polymorphisms or SNPs residing in the gene proximal promoter, intron, or 3' UTR, which is not altogether unusual. Although this could indicate that the SNPs are in LD with other phenotypically active loci, it may also be a reflection that variability in message transcription and/or turnover may explain part of the variability observed in human iris colors. Although we screened a large number of SNPs, some of the genes harbor a large number of candidate SNPs and we did not test them all. For example, the *OCA2* has ~200 known candidate SNPs in NCBI's dbSNP, and it is possible that this gene has more to teach us about variable human iris pigmentation than what we have learned from the work presented herein.

Clearly work remains to be done, objectifying the collection of iris colors from subjects, enhancing the sample size so that epistatic interactions can be explored, possibly screening other regions of the genome not screened here, and modeling the sequences that we have described to enable classification of iris colors from DNA. However, the results presented herein constitute a good first step toward solving what our results confirm is a very complex genetics problem. When this work is more fully developed, it may be possible to assign an iris color to an individual sample with reasonable certainty, and surely in this case the results herein will have some tangible value for the field of forensic science. Alternatively, as a research tool, the common haplotypes that we have identified and the complex, biologically relevant contexts within which they are found may help researchers more accurately define risk factors for pigmentation-related diseases such as cataracts and melanoma.

We thank D. C. Rao, Director of the Division of Biostatistics, Washington University, St. Louis, for help preparing this manuscript; Mark Shriver, Department of Anthropology and Human Genetics at The Pennsylvania State University for his help with the biogeographical ancestry admixture aspect of the project; and Murray Brilliant, professor of Pediatrics and Molecular and Cellular Biology at the University of Arizona for their kind advice and support of our work. We also thank Robert White for his help with sample collection. We sincerely thank the referees for their valuable suggestions for improvements on the earlier version of this article. Last, we thank the reviewers of this manuscript who suggested a number of important improvements.

LITERATURE CITED

- ABBOTT, C., I. J. JACKSON, B. CARRITT AND S. POVEY, 1991 The human homolog of the mouse brown gene maps to the short arm of chromosome 9 and extends the known region of homology with mouse chromosome 4. *Genomics* 11: 471-473.
- AKRY, J. M., H. WANG, M. XIONG, H. WU, W. LIU *et al.*, 2001 Interaction between the melanocortin-1 receptor and P genes contrib-

- utes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. *Hum. Genet.* 105: 516-520.
- BAKEMAN, G. W., N. J. MANGINI and L. H. PINTO, 1983. Discrete visual defects in pearl mutant mice. *Science* 219: 1085-1087.
- BARTON, D. E., B. S. KWON and U. FRANK, 1988. Human tyrosinase gene, mapped to chromosome 11 (q14-q21), defines second region of homology with mouse chromosome 7. *Genomics* 3: 17-24.
- BITO, L. Z., A. MATTHEY, K. J. CRUCKSHANKS, D. M. NONDAHL and O. B. CARINO, 1997. Iris color changes past early childhood. The Louisville twin study. *Arch. Ophthalmol.* 115 (5): 659-663.
- BOISSY, R. E., H. ZIAO, W. S. OETTING, L. M. AUSTIN, S. C. WILDBERG *et al.*, 1996. Mutation in and lack of expression of tyrosinase-related protein-1 (TRP-1) in melanocytes from an individual with brown oculocutaneous albinism: a new subtype of albinism classified as 'OCA3'. *Am. J. Hum. Genet.* 58: 1145-1156.
- BOX, N. F., J. R. WYETH, L. E. O'GORMAN, S. G. MARTIN and R. A. STURM, 1997. Characterization of melanocyte stimulating hormone variant alleles in twins with red hair. *Hum. Mol. Genet.* 6: 1891-1897.
- BOX, N. F., D. L. DUFFY, R. E. IRVING, A. RUSSELL, W. CHEN *et al.*, 2001. Melanocortin-1 receptor genotype is a risk factor for basal and squamous cell carcinoma. *J. Invest. Dermatol.* 116: 224-229.
- BRUKER, G. and V. P. CHOPRA, 1978. Estimation of the heritability of hair and iris color. *Anthropol. Anz.* 36 (2): 109-120.
- BRUES, A. M., 1975. Redefining human pigmentation. *Am. J. Phys. Anthropol.* 43 (3): 387-391.
- CHINTAMANNI, C. D., M. RAMSA, M.-A. COLMAN, M. F. FOX, R. T. PICKARD *et al.*, 1991. Mapping the human *CAS2* gene, the homologue of the mouse brown (b) locus, to human chromosome 9p22-pter. *Biochem. Biophys. Res. Commun.* 178: 227-235.
- COPELAND, N. G., K. W. HUTCHISON and N. A. JENKINS, 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* 33: 379-387.
- DURIHAM-PIERRE, D., J. M. GARDNER, Y. NAKATSU, R. A. KING, U. FRANK *et al.*, 1994. African origin of an intragenic deletion of the human *P* gene in tyrosinase positive oculocutaneous albinism. *Nat. Genet.* 7: 176-179.
- DURIHAM-PIERRE, D., R. A. KING, J. M. NABER, S. LAKEN and M. H. BRILLIANT, 1996. Estimation of carrier frequency of a 2.7 kb deletion allele of the *P* gene associated with OCA2 in African-Americans. *Hum. Mutat.* 7: 370-373.
- EIBERG, H. and J. MOHR, 1996. Assignment of genes coding for brown iris colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. *Eur. J. Hum. Genet.* 4 (4): 237-241.
- FLANAGAN, N., E. HEALY, A. RAY, S. PHILLIPS, C. TODD *et al.*, 2000. Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. *Hum. Mol. Genet.* 9: 2531-2537.
- FRUDAKIS, T., K. VENKATESWARLU, M. J. THOMAS, Z. GASKIN, S. GHJUPALI *et al.*, 2003. A classifier for the SNP-based inference of ancestry. *J. Forensic Sci.* 48: 771-782.
- GARDNER, J. M., Y. NAKATSU, Y. GONDO, S. LEE, M. F. LYON *et al.*, 1992. The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman syndromes. *Science* 257: 1121-1124.
- HAMABE, J., Y. FUKUSHIMA, N. HARADA, K. ABE, N. MATSUO *et al.*, 1991. Molecular study of the Prader-Willi syndrome: deletion, RFLP, and phenotype analyses of 50 patients. *Am. J. Med. Genet.* 41: 54-63.
- HANIS, C., R. CHAKRABORTY, R. FERRELL and W. SCHULL, 1986. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. *Am. J. Phys. Anthropol.* 70 (4): 433-441.
- IMESCI, P. D., I. H. WALLOW and D. M. ALBERT, 1997. The color of the human iris: a review of morphologic correlates and of some conditions that affect initial pigmentation. *Surv. Ophthalmol.* 41 (Suppl. 2): S117-S123.
- JACKSON, L. J., 1988. A cDNA encoding tyrosinase-related protein maps to the brown locus in mouse. *Proc. Natl. Acad. Sci. USA* 85 (12): 4392-4396.
- JACKSON, L. J., D. M. CHAMBERS, K. TSUKAMOTO, N. G. COPELAND, D. J. GILBERT *et al.*, 1992. A second tyrosinase-related protein, TRP-2, maps to and is mutated at the mouse *slaty* locus. *EMBO J.* 11: 527-535.
- KANETSKY, P., J. SWOYER, S. PANOSSIAN, R. HOLMES, D. GUERRY *et al.*, 2002. A polymorphism in the Agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet.* 70: 770-775.
- KING, R. A., D. TOWNSEND, W. S. OETTING and R. A. SPRITZ, 1989. An unusual pigment pattern in type I oculocutaneous albinism (OCA) resulting from a temperature-sensitive enzyme. (Abstr.) *Am. J. Hum. Genet.* 45 (Suppl.): A8.
- KLEBER, M. L., J. E. WILKINSON, J. G. GEISLER and R. P. WOYCHIK, 1995. Ectopic expression of the agouti gene in transgenic mice causes obesity, features of type II diabetes, and yellow fur. *Proc. Natl. Acad. Sci. USA* 92: 4728-4732.
- KOPPELA, S. V., L. S. ROBBINS, D. LU, E. BAACK, C. R. WHITE, JR. *et al.*, 1997. Identification of common polymorphisms in the coding sequence of the human MSH receptor (MC1R) with possible biological effects. *Hum. Mutat.* 9 (1): 30-36.
- KRUEDE, H., H. BIEBERMANN, W. LUCK, R. HORN, G. BRAANT *et al.*, 1998. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat. Genet.* 19: 155-157.
- KWON, B. S., 1993. Pigmentation genes: the tyrosinase gene family and the pmel 17 gene family. *J. Invest. Dermatol.* 100: 1345-1405.
- KWON, B. S., R. HALABAN and C. CHINTAMANNI, 1989. Molecular basis of mouse Himalayan mutation. *Biochem. Biophys. Res. Commun.* 161: 252-260.
- KWON, B. S., C. CHINTAMANNI, C. A. KOZAK, N. G. COPELAND, D. J. GILBERT *et al.*, 1991. A melanocyte-specific gene, Pmel-17, maps near the silver coat color locus on mouse chromosome 10 and is in a syntenic region on human chromosome 12. *Proc. Natl. Acad. Sci. USA* 88: 9228-9232.
- KWON, H. Y., S. J. BULTMAN, C. LOFFLER, W.-J. CHEN, P. J. FURDON *et al.*, 1994. Molecular structure and chromosomal mapping of the human homolog of the agouti gene. *Proc. Natl. Acad. Sci. USA* 91: 9760-9764.
- LEE, S.-T., R. D. NICHOLLS, R. E. SCHNUR, L. C. GUIDA, J. LU-KUO *et al.*, 1994. Diverse mutations of the *P* gene among African-Americans with type II (tyrosinase-positive) oculocutaneous albinism (OCA2). *Hum. Mol. Genet.* 3: 2047-2051.
- LINDSEY, J. D., H. L. JONES, E. G. HEWITT, M. ANGERT and R. N. WEINER, 2001. Induction of tyrosinase gene transcription in human iris organ cultures exposed to latanoprost. *Arch. Ophthalmol.* 119 (6): 853-860.
- LLOYD, V., M. RAMASWAMI and H. KRAMER, 1998. Not just pretty eyes: Drosophila eye-colour mutations and lysosomal delivery. *Trends Cell Biol.* 8 (7): 257-259.
- LYON, M. F., T. R. KING, Y. GONDO, J. M. GARDNER, Y. NAKATSU *et al.*, 1992. Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (p) locus of the mouse. *Proc. Natl. Acad. Sci. USA* 89: 6968-6972.
- NEWTON, J. M., O. COHEN-BARAK, N. HAGIWARA, J. M. GARDNER, M. T. DAVENSON *et al.*, 2001. Mutations in the human orthologue of the mouse underwhite gene (*uw*) underlie a new form of oculocutaneous albinism, OCA4. *Am. J. Hum. Genet.* 68: 981-988.
- OETTING, W. S. and R. A. KING, 1991. Mutations within the promoter region of the tyrosinase gene in type I (tyrosinase-related) oculocutaneous albinism. (Abstr.) *Clin. Res.* 39: 267A.
- OETTING, W. S. and R. A. KING, 1992. Molecular analysis of type I-A (tyrosine negative) oculocutaneous albinism. *Hum. Genet.* 90: 258-262.
- OETTING, W. S. and R. A. KING, 1993. Molecular basis of type I (tyrosinase-related) oculocutaneous albinism: mutations and polymorphisms of the human tyrosinase gene. *Hum. Mutat.* 2: 1-6.
- OETTING, W. S. and R. A. KING, 1995. Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. *Hum. Mutat.* 15: 99-115.
- OOI, C. E.-J., E. MOREIRA, E. C. DELA-ANGELICA, G. POY, D. A. WASSARMAN *et al.*, 1997. Altered expression of a novel adaptin leads to defective pigment granule biogenesis in the Drosophila iris color mutant garnet. *EMBO J.* 16 (15): 4508-4518.
- REBBECK, T. R., P. A. KANETSKY, A. H. WALKER, R. HOLMES, A. C. HALPERN *et al.*, 2002. *P* gene as an inherited biomarker of human eye color. *Cancer Epidemiol. Biomarkers Prev.* 11 (8): 782-784.
- ROBBINS, L. S., J. H. NADEAU, K. R. JOHNSON, M. A. KELLY, L. ROSELLI-REHFS *et al.*, 1993. Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* 72: 827-834.
- SCHIOTH, H. B., S. R. PHILLIPS, R. RUDZIS, M. A. BIRCH-MACHIN,

- J. E. WIKBERG *et al.*, 1999 Loss of function mutations of the human melanocortin 1 receptor are common and are associated with red hair. *Biochem. Biophys. Res. Commun.* **260**: 488-491.
- SCHMIDT A., and F. BEERMANN, 1994 Molecular basis of dark-eyed albinism in the mouse. *Proc. Natl. Acad. Sci. USA* **91** (11): 4756-4760.
- SHRIVER, M., E. PARRA, S. DIOS, C. BONILLA, H. NORTON *et al.*, 2003 Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum. Genet.* **112**: 387-399.
- SMITH, R., E. HEALY, S. SIDDIQUI, N. FLANAGAN, P. M. STEIJLEN *et al.*, 1998 Melanocortin 1 receptor variants in an Irish population. *J. Invest. Dermatol.* **111**: 119-122.
- SPRITZ, R. A., 1995 A study in scarlet. *Nat. Genet.* **11**: 225-226.
- STEENLAND, K., I. BRAY, S. GREENLAND and P. BOFFETTA, 2000 Empirical Bayes adjustments for multiple results in hypothesis-generating or surveillance studies. *Cancer Epidemiol.* **9**: 895-903.
- STEPHENS, M., N. SMITH and P. DONNELLY, 2001 A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**: 978-989.
- STROBEL, M. G., P. K. SEPERACK, N. G. COPELAND and N. A. JENKINS, 1990 Molecular analysis of two mouse dilute locus deletion mutations: spontaneous dilute lethal-20J and radiation-induced dilute prenatal lethal Aa2 alleles. *Mol. Cell. Biol.* **10**: 501-509.
- STURM, R., R. TEASDALE and N. BOX, 2001 Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene* **277**: 49-62.
- VALVERDE, P., E. HEALY, I. JACKSON, J. L. REES and A. J. THODY, 1995 Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* **11**: 328-330.
- WILSON, S. M., R. YIP, D. A. SWING, T. N. O'SULLIVAN, Y. ZHANG *et al.*, 2000 A mutation in *Rab27a* causes the vesicle transport defects observed in ashen mice. *Proc. Natl. Acad. Sci. USA* **97**: 7933-7938.
- ZAYKIN, D., L. ZHIVOTOVSKY and B. S. WEIR, 1995 Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* **96** (1-2): 169-178.

Communicating editor: P. J. OEFNER